Journal of Neural Engineering

HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification

IOP Publishing Journal Title
Journal **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

HS-CNN: A CNN with Hybrid Convolution Scale for EEG Motor Imagery Classification

Guanghai Dai, Jun Zhou*, Jiahui Huang, Ning Wang

School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, P. R. China

E-mail: zhouj@uestc.edu.cn

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Abstract

Objective. The EEG motor imagery classification has been widely used in healthcare applications such as mobile assistive robots and post-stroke rehabilitation. Recently, CNN-based EEG motor imagery classification methods have been proposed and achieve relatively high classification accuracy. However, these methods use single convolution scale in the CNN, while the best convolution scale differs from subject to subject. This limits the classification accuracy. Another issue is that the classification accuracy degrades when the training data is limited. *Approach*. To address these issues, we have proposed a hybrid-scale CNN architecture with a data augmentation method for EEG motor imagery classification. *Main results.* Compared with several state-of-the-art methods, the proposed method achieves an average classification accuracy of 91.57% and 87.6% on two commonly used datasets, which outperforms several state-of-the-art EEG motor imagery classification methods. *Significance.* The proposed method effectively addressed the issues of existing CNN-based EEG motor imager classifcation methods and improved the classification accuracy.

Keywords: EEG, Motor Imagery, CNN

1. Introduction

Brain-Computer Interface (BCI) is to enable communication between a brain and a device. As no peripheral nerve and muscle are involved in the process, it allows those with neuromotor disorders, nervous system injuries, or limbs amputation to control machine by their brains. Based on this, several mobile assistive robots have been developed, e.g., BCI-based wheelchair [1,2], upper (or lower) limb assistive-robot [3–6], and so on. These mobile assistive robots collect brain signals of the patients and convert them to commands. This helps patients with nervous system injuries or limbs amputation live independently. Also, BCI has been a new tool for post-stroke rehabilitation [7–9]. In traditional post-stroke rehabilitation approaches, active motor training and pharmacological interventions are used for the motor recovery of stroke patients [10]. However, physical movements are often difficult to the stroke patients [11]. Unlike Active Motor Training, BCI-based post-stroke rehabilitation systems help stroke patients do exercises through their brain signals rather than muscles. Recently, BCI has also been applied to gaming, e.g., "Tetris" [12], "Pacman" [13], "Pinball" [14] and "World of Warcraft" ("WoW") [15], and so on.

The sources of BCI signal generally fall into two categories: one is based on neuron discharge activity, such as magnetoencephalography (MEG), electroencephalography (EEG), electrocorticogram (ECoG), action potentials and local field potentials (LFPs), the other is based on neuronal metabolic activity, that is, the observation of changes in oxyhemoglobin and deoxygenated hemoglobin, such as functional magnetic resonance and near-infrared spectral imaging. Among them, EEG is one of the most widely used

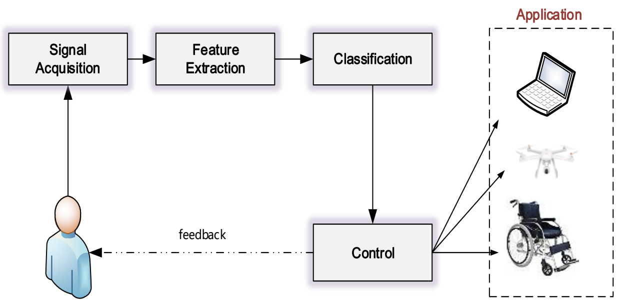

Figure 1. Motor imagery classification.

BCI signal sources because of its non-invasiveness and high temporal resolution [16].

Motor imagery classification is one of the most popular EEG-based BCI paradigms. The user can generate induced activity in the motor cortex area of his/her brain by imagining motor movements (e.g., left/right-hand movement) and the resultant EEG signals can be classified automatically. EEG motor imagery classification generally consists of four parts: signal acquisition, feature extraction, classification and control, as shown in Figure 1. Feature extraction plays an important role in EEG motor imagery classification. Most existing feature extraction methods are highly dependent on manually designed features based on human knowledge and experience [17–21]. However, due to the limitation of human knowledge and experience, manually designed features may lead to limited classification accuracy [22,23]. Also, finding effective feature usually requires extensive experiments and observations, which consumes a large amount of time and effort. Recently some researchers found that it is possible to use neural networks to perform automatic feature extraction and classification [24–30]. This approach avoids the limitation of human knowledge and experience, and is promising in achieving higher accuracy. It also reduces the time and effort of manual feature design, as effective features can be automatically generated through network training.

However, the existing CNN-based classification methods use single convolution kernel size which limits the classification accuracy (will be discussed in details in later sections). They also suffer from accuracy issue when available training data is limited. By considering the above issues, in this paper, a hybrid-scale CNN (HS-CNN) classification method and a data augmentation method are proposed for the EEG motor imagery classification to improve the accuracy. The rest of the paper is organized as follows: Section 2 reviews the work related to the classification of EEG motor imagery. Section 3 describes the proposed HS-CNN method and data augmentation method. Section 4 gives the experimental results and discussion, and Section 5 concludes the paper.

2.Related Works

For EEG signals, motion-related information is usually contained in the time and frequency domain. In addition, the position of electrodes used to record the EEG signals also affects the recognition of motion-related information. Therefore, the feature extraction and classification of EEG signals are usually performed in the time, frequency and space domains.

For time-frequency feature extraction, there are many works that use STFT [18,30] or wavelet [19–21,31] to extract the time-frequency features of EEG signals, but they generally extract time-frequency features at a fixed time period or in the same frequency band, but the time period and frequency band containing motion-related information may vary from subject to subject, or from time to time for the same subject. Therefore, it is not able to achieve a high classification accuracy for different subjects. [32] uses Wavelet Packet Decomposition (WPD) to extract the time-frequency features and then uses Dynamic frequency feature selection (DFFS) algorithm to select the most accurate feature for each subject; [33] first selects the time period that is most related to the motion information by ERS/ERD phenomenon, then uses WPD to extract the time-frequency features, and then adopts feature selection. Although [32,33] have improved the classification accuracy, it takes a large amount of time to select the most suitable features for each subject, and the method is not generalized.

In terms of space domain, Common Spatial Pattern (CSP) is one of the effective and popular feature extraction methods in EEG signal processing [6,33,34], but the performance of CSP depends on the selected frequency band. To address this issue, [35] proposed Filter Bank Common Spatial Patter (FBCSP), an improved version of CSP. FBCSP applies CSP to different frequency bands, and then uses a feature selection algorithm to automatically select subject-specific features, which greatly improves the classification accuracy. However, the FBCSP uses a fixed time period, ignoring the difference between the best time periods of different subjects [17], and does not make full use of time domain information. Although some work has added the best time choice based on FBCSP, the accuracy improvement is not significant [36,37].

In recent years, deep learning has been applied in many fields such as image and speech recognition, and achieved great success. There are also some studies that apply these approaches to motor imagery classification, since deep learning methods can automatically extract EEG signal features and avoid the limitations of manually designed features. Among these approaches, CNN is the most popular for its great success in the field of computer vision. [26] proposed a CNN with shallow and deep architecture, and applied some recently developed techniques such as batch normalization and exponential linear unit activation functions to obtain a higher accuracy than FBCSP. A parallel CNN with

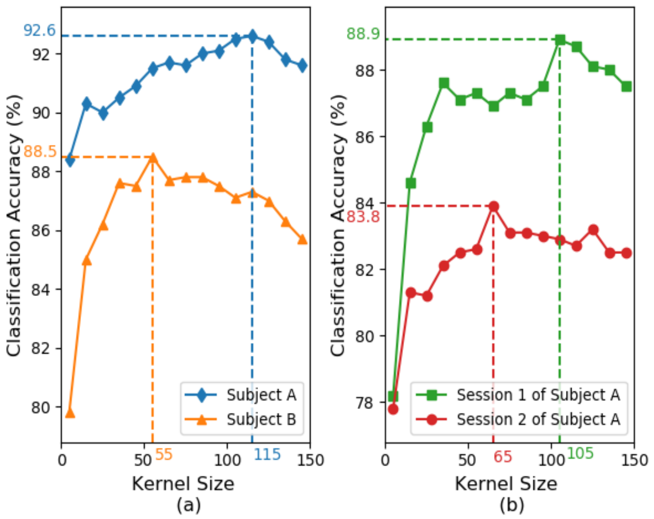

Figure 2. (a) Classification accuracy of two different test subjects plotted against the kernel size used. (b) Classification accuracy of two different sessions of the test subject A plotted against the kernel size used.

Multilayer perceptron (MLP) was proposed by [38] to process static energy (trail energy) and dynamic energy (envelope energy) separately and achieve a high classification accuracy. [27] proposed a CNN which can be applied to a variety of popular BCI paradigms such as motor imagery and P300 visual-evoked potentials, and obtained comparably high accuracy to the state-of-the-art methods. A deep learning method combining CNN and SAE was proposed by [30], in which they use CNN to extract features and then use deep network SAE to perform classification. In addition to CNN, there are some studies using other neural network topologies. [25] proposed a Deep Belief Network (DBN) for motor imagery EEG classification, which was trained on raw EEG data and obtained a 4-6% improvement in the left- and right-hand motor imagery classification. [29] developed a deep learning scheme named FDBN, which is based on restricted Boltzmann machine (RBM) and uses frequency domain features as input. It achieves 5% accuracy improvement compared with the top three methods in BCI competition IV. There are also some studies [39,40] using Spiking Neural Network (SNN) for motor imagery classification as SNN raises the level of biological realism by utilizing spikes [41,42].

The existing CNN-based methods still face two issues. Firstly, many methods have applied CNN to EEG motor imagery classification [24,26,27,43]. However, the CNNs proposed in the existing work uses a single convolution scale to extract the EEG features. This limits the classification accuracy as the best scale may differ significantly from subject to subject, or from time to time for the same subject. Second, a common issue of CNN is that it requires a large amount of training data to achieve high classification accuracy, but it is very difficult to obtain training samples for motor imagery classification as the sampling process takes a lot of procedures and requires very strict rules (e.g. the eye and muscle movement are not allowed, and the test subject needs to have a great deal of concentration during the sampling period).

3.Method

To address the above issues, we have proposed the HS-CNN, which is able to perform convolution in a hybrid scale to improve the classification accuracy. We have also proposed a data augmentation method to generate artificial training data based on the real training data to further improve the accuracy. The details of the proposed methods are described below.

3.1 Proposed HS-CNN Structure

CNN is a feedforward network, and its structure is inspired by the visual cortex of human brain. CNN generally consists of alternating convolution and pooling layers and fully connected layers at the end. The convolution layers and the pooling layers are designed to extract features, while the fully connected layers are used for classification. The CNN network mainly mimics three important characteristics of the cerebral cortex: local connectivity, invariance to location, and invariance to local transition [44,45]. It is an attractive neural network structure as it has fewer parameters and is less prone to overfitting than MLP.

In the field of computer vision, CNN is designed to extract edges, corners and textures of 2D/3D images. During the design, one of main considerations is to find a CNN architecture with appropriate convolution scale (i.e. kernel size) to well extract the features of images for achieving a high classification accuracy. However, for motor imagery classification, the best kernel size varies from subject to subject. For example, we have investigated the impact of kernel size on different test subjects. Figure 2 shows the classification accuracy of two different test subjects with different kernel sizes. It can be seen that the best kernel size for subject A is 1×115 while the best kernel size for subject B is 1×55. Also, for the same subject A, the best kernel size for session 1 and session 2 are 1×105 and 1×65 respectively. They indicate that the best kernel size varies from subject to subject, and for the same subject the best kernel size varies from time to time. Inspired by this finding, we have proposed a CNN with hybrid convolution scale (named as HS-CNN) to improve the classification accuracy.

Figure 3 shows the architecture of the proposed HS-CNN. The earlier work [46–48] has already shown that the motor imagery tasks will result in desynchronization (ERD) and synchronization (ERS) phenomena in the mu and beta frequency bands (8-32 Hz). In the recent years, some work has shown that the theta band (4-7 Hz) also plays a crucial role in motor imagery tasks [49,50] as theta band ERD significantly differs between the left/right-hand motor imagery tasks. Therefore, we selected the theta frequency band (4-7 Hz), mu frequency bands (8-13 Hz) and beta frequency bands (13-32 Hz) for the signal filtering to acquire sufficient information for the subsequent classification. Therefore, we designed a filter

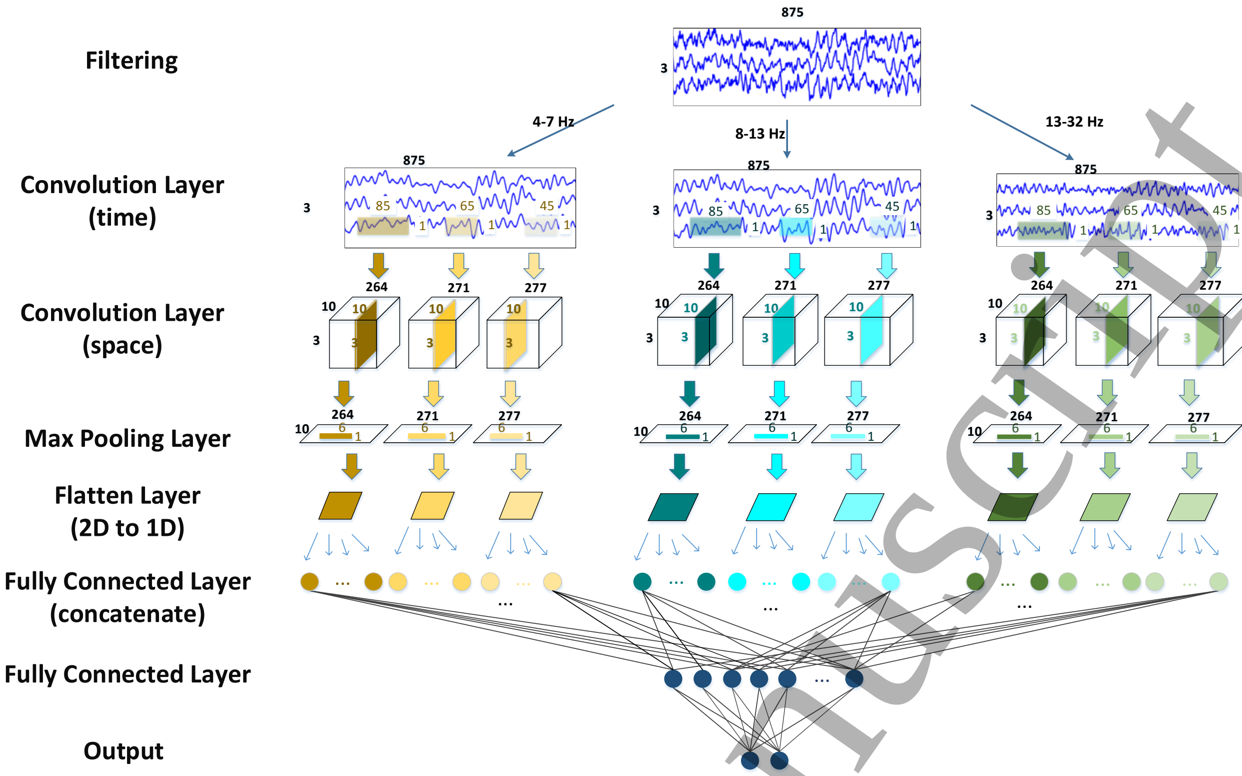

Figure 3. Proposed HS-CNN architecture.

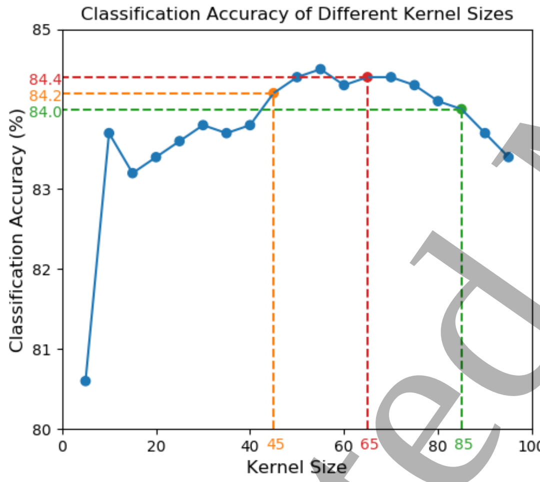

Figure 4. Average classification accuracy across the tested subjects with different kernel sizes.

bank of 4-7 Hz, 8-13 Hz, and 13-32 Hz to obtain EEG information in these frequency bands. In addition, we proposed a CNN with hybrid convolution scale (named as HS-CNN) as we found that the best kernel size varies from subject to subject, and for the same subject the best kernel size varies from time to time, as shown in Figure 2. Regarding the convolution kernel size, before the selection of the kernel size, we have performed extensive experiments to investigate the impact of kernel size on the average classification accuracy across the tested subjects. The results are shown in the Figure 4. As can be seen from the figure, the classification accuracy varies with the kernel size. According to the results, we have selected three different kernel sizes (1×45, 1×65 and 1×85) with relatively high accuracy (>84%). Also, the three kernel sizes are distributed with some distance for suiting different subjects. Therefore, we apply three convolution kernel sizes (i.e. 1×45, 1×65 and 1×85) in each frequency bands to extract the time-domain features of the EEG signals with different convolution scales. Then, for each kernel-size branch, a one-dimensional kernel along the vertical axis is used to extract the space-domain information. In this way, EEG information in different domains (time, frequency and space) with different convolution scales is extracted. For the feature fusion, in the Flatten layer, we map the 2D features to 1D features by concatenating the rows in the 2D feature into a 1D vector. This is done for all the nine branches. Then, after Flatten layer, we concatenate all the 1D features from the nine branches into a 1D vector and use the vector as the input of the fully connected layer. The proposed network uses Exponential Linear Units (ELUs) proposed in [51] as the activation function as it is able to speed up learning in neural networks and improve the classification accuracy. To prevent overfitting, L2 regularization and dropout technique are used. The detailed parameters of the proposed HS-CNN architecture are given in Table 1.

3.2 Data Augmentation

For neural network, the classification accuracy is highly dependent on the amount of training data. When the amount

Table 1. Architecture of proposed CNN

Layer Type	Number of Filters	Size of Feature Map	Size of Kernel	Number of Stride	Parameters
Input layer		3(height)×875(width) ×1(channel)			
1st convolutional layer (4-9/8-15/14-31 Hz)					
Branch1	10	3×264×10	1×85	1×3	850
Branch2	10	3×271×10	1×65	1×3	650
Branch3	10	3×277×10	1×45	1×3	450
2nd convolutional layer (4-9/8-15/14-31 Hz)					
Branch1	10	1×264×10	3×1	1×1	300
Branch2	10	1×271×10	3×1	1×1	300
Branch3	10	1×277×10	3×1	1×1	300
Max pooling layer (4-9/8-15/14-31 Hz)					
Branch1	10	1×44×10	1×6	1×6	
Branch2	10	1×46×10	1×6	1×6	
Branch3	10	1×47×10	1×6	1×6	
1st full connected layer (with L2 regularization)		100			405000
Dropout layer					
2st full connected layer		2			200

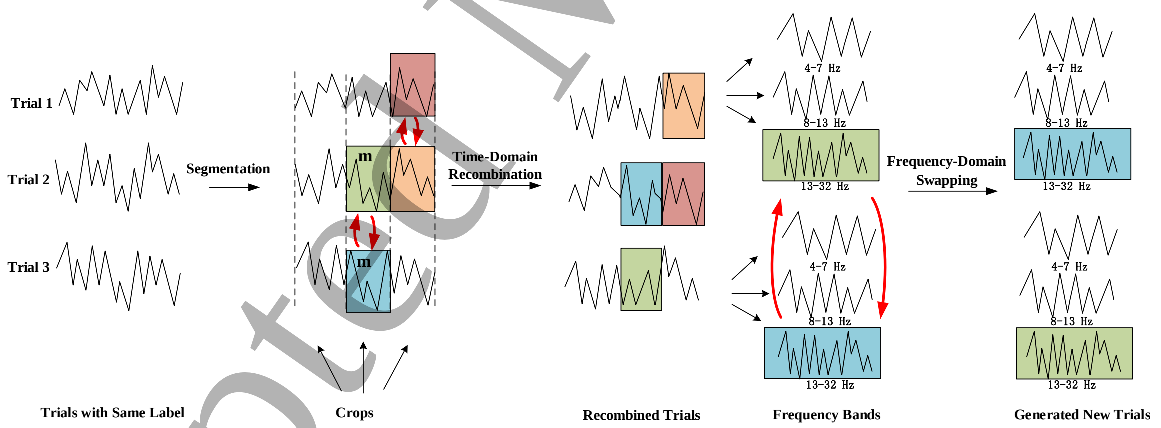

Figure 5. Proposed data augmentation method.

of training data is limited, the classification accuracy is usually low. For motor imagery classification, it is difficult to acquire training data as each experiment has complex procedure and strict rules to follow. For example, eye movements and muscle movements can seriously affect EEG signals and are therefore not allowed in experiments. Also, the test subject is required to fully focus when imagining the left/right-hand movement, which becomes difficult as the number of experiments increases.

To address this issue, two types of data augmentation methods have been proposed in the past [52–56]. One of them is to perform data transformation to generate more training data from the original training data [52–54]. The other is to add noise to the original training data to generate new training data [55,56]. However, the improvement on classification accuracy is limited. In this work, we have proposed a new data augmentation method to generate artificial training data based on the real training data to improve the classification accuracy.

Table 2. Architecture of the baseline CNN

Layer Type	Number of Filters	Size of Feature Map	Size of Kernel	Number of Stride	Parameters
Input layer		3(height)×875(width) × 1(channel)			
1st convolutional layer	90	3×277×90	1×45	1×3	4050
2nd convolutional layer	90	1×277×90	3×1	1×1	2700
Max pooling layer	90	1×47×90	1×6	1×6	
1st full connected layer (with L2 regularization)		100			423000
Dropout layer					
2st full connected layer		2			200

Table 3. Wilcoxon signed-rank test between proposed HS-CNN and other methods

Models	Baseline	Baseline65	Baseline85	EEGNet	DeepNet	ShallowNet
P-values	0.0382	0.0284	0.0209	0.0382	0.0284	0.0284

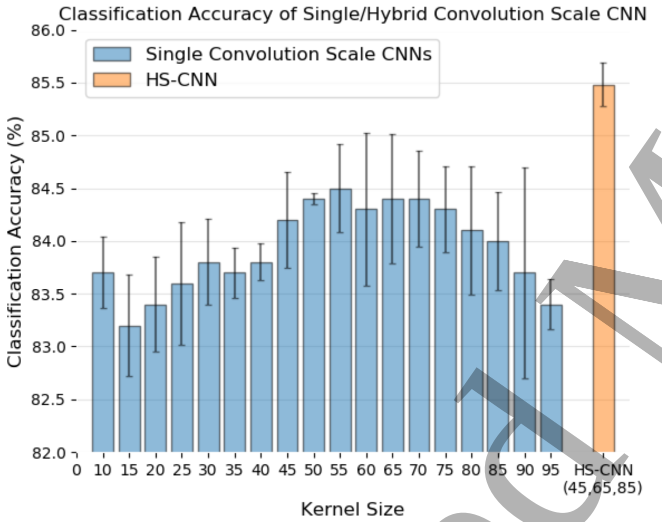

Figure 6. Experimental comparison between proposed HS-CNN and CNNs with single convolution scale.

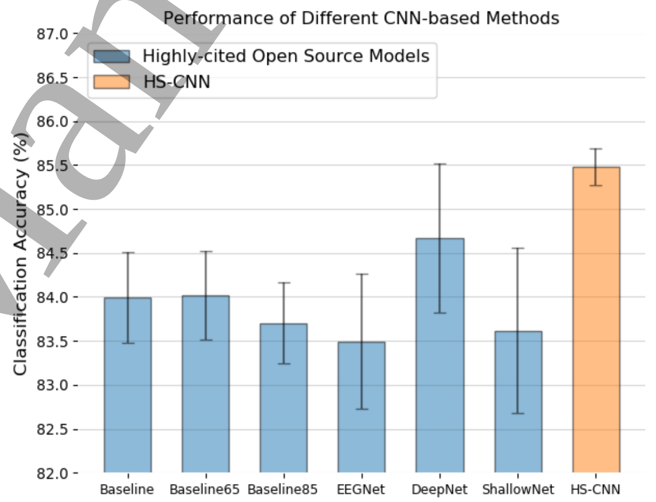

Figure 7. Comparison of classification accuracy on the BCI Competition IV 2b dataset.

The details of the proposed data augmentation method are described below:

Our proposed data augmentation method can be divided into 3 stages, as shown in Figure 5.

1) Segmentation: We first perform segmentation on the the input trials (i.e. left/right-hand motor imageries) with the same label. Each trial is segmented into 3 crops. As the length of a trial is 875 which cannot be divided exactly, the first two crops each contain 292 points and the 3rd crop contains 291 points. As shown in (1) and (2), the ith crop in the j^{th} trial is represented by L_i^j/R_i^j, where L represents left-hand movement and R represents right-hand movement:

$$L^j = [L_1^j, L_2^j, \cdots, L_n^j] \tag{1}$$

$$R^j = [R_1^j, R_2^j, \cdots, R_n^j] \tag{2}$$

2) Time-Domain Recombination: The crops with the same labels are then recombined to generate new trials. For the same person and the same class, the crops at the same position from multiple trials are randomly swapped and recombined to generate recombined trials. For example, the m^{th} crop in the p^{th} trial can be swapped with the m^{th} crop in the q^{th} trial.

3) Frequency-Domain Swapping: After the band-pass

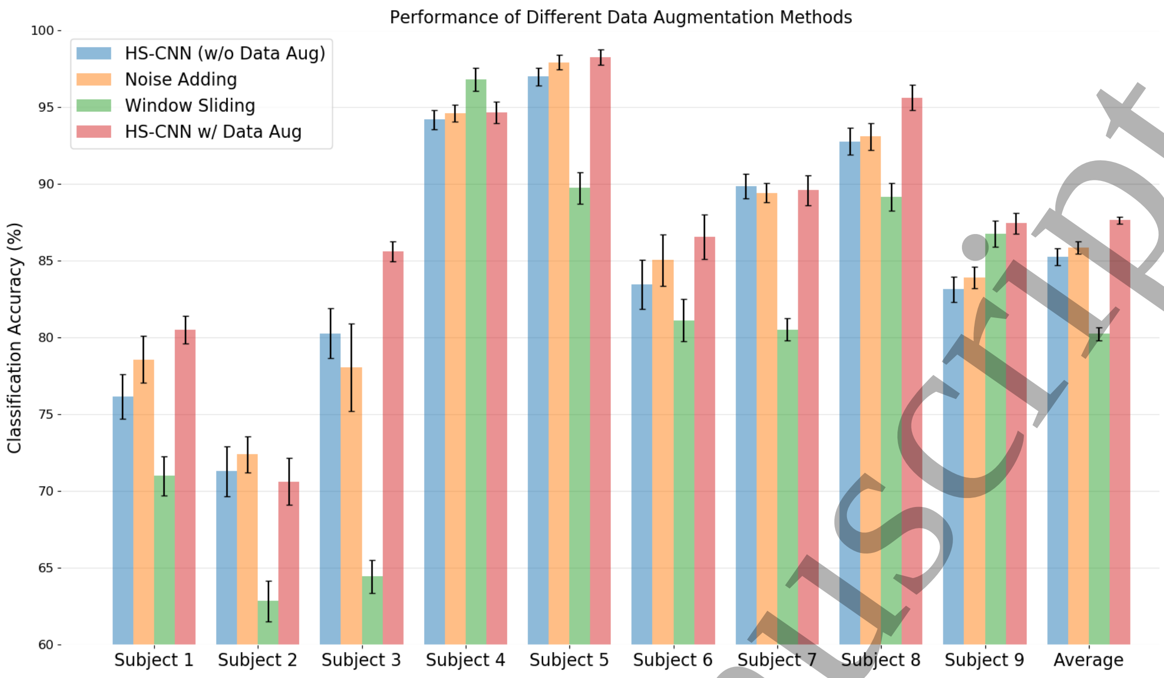

Figure 8. Experimental results on the BCI Competition IV 2b dataset using different data augmentation techniques

Table 4. Wilcoxon signed-rank test between proposed data augmentation method and other data augmentation methods

Models	Noise Adding	Window Sliding
P-values	0.0218	0.0093

filtering, the filtered trials (from the same person and the same class) of the same frequency band are randomly swapped to form new filtered trials. For example, as shown in Figure 5, the 3rd frequency band of the x th trial can be swapped with the 3rd frequency band of the y th trial.

In this work, we repeat 2) and 3) for multiple times to generate 3 times training data as original.

4.Experimental Results

4.1 Dataset and Experimental Method

In this work, BCI Competition IV 2a and 2b datasets [57] are used to evaluate the proposed method. BCI Competition IV 2a collects EEG data from nine healthy subjects. For each subject, 2 sessions of data are collected. Each session has 288 trials. The total trial number is 5184. BCI Competition IV 2b collects EEG data from another nine healthy subjects. For each subject, 5 sessions of data are collected. Each of the first 2 sessions has 120 trials and each of the last 3 sessions has 160 trials. The total trial number is 6520. Two types of trials are included in these datasets: left-hand movement and right-hand movement. The EEG data is collected by three electrodes (C3, CZ, C4) and the electrodes are placed following the international 10-20 system [58]. The sampling frequency is 250 Hz, and the time period of a trial ranges from 3.5s to 7s. So a trial is a 3×875 matrix.

In our experiments, the dropout probability is set to 0.8 and the L2 regularization parameter is set to 0.01. Stochastic gradient descent (SGD) is used as our optimization method. The initial learning rate is set to 0.1 and learning rate decays every 10 epochs (training for 400 epochs) with an exponential decay rate of 0.9.

4.2 Performance of the proposed HS-CNN

To evaluate the performance of proposed HS-CNN, comparative experiments have been conducted between the HS-CNN and CNNs with single convolution scale. The baseline CNN is designed following a frequently cited method for motor imagery classification [26]. This method has a relatively high classification accuracy among the state-of-the-art methods. The architecture of the baseline CNN is shown in Table 2. As the original CNN architecture in [26] has 44 input channels and 4 classified outputs, we have slightly modified it to suit our case (i.e. 3 input channels and 2 classified outputs). For fair comparison, the proposed HS-CNN uses almost the same number of parameters (slightly less) as the baseline CNN, as shown in Table 1 and 2. To evaluate the impact of kernel size on the classification accuracy, we have also varied the kernel size of the 1st convolution layer in the baseline CNN from 1×45 to other sizes (different sizes between 1×10 and 1×95). The results are shown in Figure 6. Compared with the CNNs with single convolution scale, the proposed HS-CNN with hybrid convolutional scale has much higher accuracy. In

Table 5. Compare table of the proposed method with state-of-the-art methods

	Shahid et al 2011* [59]	Ang et al 2012* [60]	Saa et al 2012* [34]	Luo et al 2016 [32]	Li et al 2017 [33]	Lu et al 2017 [29]	Tabar et al 2017 [30]	Zheng et al 2018 [61]	Lotte & Guan, 2011[62]	Raza et al., 2016 [63]	Raza et al., 2016 [63]	Gaur et al 2015 [64]	Gaur et al 2018 [65]	Proposed
Dataset	2b	2b	2b	2a/2b	2b	2b	2b	2b	2a	2a/2b	2a/2b	2a/2b	2a	2a/2b
S1	77.0	70.0	80.0	63.69/ 73.2	84.6	81.0	76.0	72.5	88.89	90.28/ 71.9	90.28/ 70.3	66.7/ 62.8	91.49	90.07/ 80.5
S2	64.5	60.0	66.0	61.97/ 67.5	66.3	65.0	65.8	56.4	51.39	54.17/ 50.0	57.64/ 50.6	63.9/ 67.1	60.56	80.28/ 70.6
S3	61.0	61.0	53.0	91.09/ 63	62.9	66.0	75.3	55.6	96.53	93.75/ 52.8	95.14/ 52.8	77.8/ 98.7	94.16	97.08/ 85.6
S4	96.5	97.5	98.5	61.72/ 97.4	95.8	98.0	95.3	97.2	70.14	64.58/ 93.4	65.97/ 93.8	63.2/ 88.4	76.72	89.66/ 94.6
S5	82.0	92.8	93.5	63.41/ 95.5	89.2	93.0	83.0	88.4	54.86	57.64/ 54.4	61.11/ 63.8	72.2/ 96.3	58.52	97.04/ 98.3
S6	84.5	81.0	89.0	66.11/ 86.7	97.9	88.0	79.5	78.7	71.53	65.28/ 73.1	65.28/ 74.1	70. 1/ 75.3	68.52	87.04/ 86.6
S7	75.0	77.5	81.5	59.57/ 84.7	82.1	82.0	74.5	77.5	81.25	65.2/ 62.2	61.11/ 61.9	64.6/ 72.2	78.57	92.14/ 89.6
S8	91.0	92.5	94.0	62.84/ 95.9	86.3	94.0	75.3	91.9	93.75	90.97/ 77.8	91.67/ 83.1	76.4/ 87.8	97.01	98.51/ 95.6
S9	87.0	87.2	90.5	84.46/ 92.6	97.1	91.0	73.3	83.4	93.75	85.42/ 75	86.11/ 77.2	77. 1/ 85.3	93.85	92.31/ 87.4
AVG	80.0	80.0	83.0	68.32/ 84.1	84.7	84.0	77.6	78.0	78.01	73.84/ 67.9	74.92/ 69.7	70.2/ 81.6	79.93	91.57/ 87.6

The * sign indicates that the classification accuracy of the item is calculated based on kappa value; AVG represents the average classification accuracy of nine subjects. The compared methods all utilize the same data set (2008 BCI Competition IV 2a or/and 2008 BCI Competition IV 2b).

addition, we have retrained 3 highly-cited open source models (EEGNet, DeepNet, ShallowNet) [26,27] and evaluated their performance using the BCI Competition IV 2b dataset. The comparison results are given in Figure 7, where 'baseline' 'baseline65' and 'baseline85' refer to the CNNs with single kernel size of 1×45, 1×65 and 1×85, respectively. The experiments are repeated and the average accuracy with error bar are plotted. As can be seen from the figure, the proposed HS-CNN achieves an average accuracy of 85.6%, which is higher than the other methods. To further validate this result, Wilcoxon signed-rank test has been conducted for all the methods and the obtained p-values are given in Table 3. It can be seen from the table that all the p-values are less than 0.05 indicating that the proposed HS-CNN outperform other compared methods statistically.

4.3 Performance of the proposed data augmentation method

We have also evaluated the performance of the proposed data augmentation method and compared it with the existing data augmentation methods including window sliding [26,52,53] and noise adding [55,56]. For fair comparison, all of these data augmentation methods are applied to the proposed HS-CNN. In the window sliding method, the new training trials are generated by sliding a window of 500 points with a step of 100 points, as suggested in [26]. In the noise adding method, the new training trials are generated by adding Gaussian white noise to original trials, as suggested in [55]. For the proposed data augmentation method, each trial is segmented into 3 crops for time-domain recombination and frequency-domain swapping, as described in Section 3.2.

The comparison results are shown in Figure 8. The HS-CNN without data augmentation has an average classification accuracy of 85.6%. The accuracy becomes 80.1% and 86.1% after using the windows sliding and noise adding methods, respectively. The reason why the accuracy even decreases after using the window sliding is probably due to the loss of motion-related information after changing the window size from 875 to 500. With the proposed data augmentation method, the accuracy is improved to 87.6%, which is the highest among the compared methods. In addition, a Wilcoxon signed-rank test is applied to further validate the results and the obtained p-values are given in Table 4. It can be seen from the table that all the p-values are less than 0.05 indicating that the proposed data augmentation methods have better performance than other compared methods statistically.

4.4 Comparison of Proposed Method with State-of-the-Art Methods

Table 5 compares the performance of the proposed method with that of several state-of-the-art methods based on 2008 BCI Competition IV 2a and 2b datasets. It can be seen that the proposed method has the highest average classification accuracy among the compared methods for both datasets. More specifically, the improvement in the average classification accuracy is up to 23.25% for the 2a dataset and up to 19.7% for the 2b dataset. For the 2a dataset, compared with other methods, the proposed method shows a large improvement in the accuracy of subject 5 (up to 39.4% increase) and subject 7 (up to 32.57% increase). For the 2b dataset, the proposed method shows a large improvement in the accuracy of subject 2 (up to 20.6% increase), and subject 7 (up to 27.7% accuracy increase).

5.Conclusion

Existing CNN-based EEG motor imagery classification methods use single convolution scale in the CNN layer to extract the EEG features. This results in limited classification accuracy as the best convolution scale may differ from subject to subject, or from time to time for the same subject. Another issue is the classification accuracy is usually low with limited training data. In this work, we have proposed a hybrid-scale CNN architecture with a data augmentation method to improve the accuracy of EEG motor imagery classification. For the 2008 BCI Competition IV 2a dataset, the proposed EEG motor imagery classification method achieves an average classification accuracy of 91.57%. Compared with several state-of-the-art methods, the improvement is from 11.64% to 23.25%. For the 2008 BCI Competition IV 2b dataset, the proposed method achieves an average classification accuracy of 87.6%, and the improvement is from 2.9% to 19.7%.

References

[1] Tsui C S L, Gan J Q and Hu H 2011 A Self-Paced Motor Imagery Based Brain-Computer Interface for Robotic Wheelchair Control *Clin EEG Neurosci* **42** 225–9

[2] Ng D W, Soh Y and Goh S 2014 Development of an Autonomous BCI Wheelchair *2014 IEEE Symposium on Computational Intelligence in Brain Computer Interfaces (CIBCI)* 2014 IEEE Symposium on Computational Intelligence in Brain Computer Interfaces (CIBCI) pp 1–4

[3] He Y, Eguren D, Azorín J M, Grossman R G, Luu T P and Contreras-Vidal J L 2018 Brain–machine interfaces for controlling lower-limb powered robotic systems *J. Neural Eng.* **15** 021004

[4] Bhagat N A, Venkatakrishnan A, Abibullaev B, Artz E J, Yozbatiran N, Blank A A, French J, Karmonik C, Grossman R G, O'Malley M K, Francisco G E and Contreras-Vidal J L 2016 Design and Optimization of an EEG-Based Brain Machine Interface (BMI) to an Upper-Limb Exoskeleton for Stroke Survivors *Front Neurosci* **10** 122

[5] Riener R 2016 The Cybathlon promotes the development of assistive technology for people with physical disabilities *J Neuroeng Rehabil* **13**

[6] Meng J, Zhang S, Bekyo A, Olsoe J, Baxter B and He B 2016 Noninvasive Electroencephalogram Based Control of a Robotic Arm for Reach and Grasp Tasks *Scientific Reports* **6** 38565

[7] Tung S W, Guan C, Ang K K, Phua K S, Wang C, Zhao L, Teo W P and Chew E 2013 Motor imagery BCI for upper limb stroke rehabilitation: An evaluation of the EEG recordings using coherence analysis *Conf Proc IEEE Eng Med Biol Soc* **2013** 261–4

[8] Ang K K, Guan C, Phua K S, Wang C, Teh I, Chen C W and Chew E 2012 Transcranial direct current stimulation and EEG-based motor imagery BCI for upper limb stroke rehabilitation *Conf Proc IEEE Eng Med Biol Soc* **2012** 4128–31

[9] Johnson N N, Carey J, Edelman B J, Doud A, Grande A, Lakshminarayan K and He B 2018 Combined rTMS and virtual reality brain–computer interface training for motor recovery after stroke *J. Neural Eng.* **15** 016009

[10] Calautti C and Baron J-C 2003 Functional neuroimaging studies of motor recovery after stroke in adults: a review *Stroke* **34** 1553–66

[11] Ang K K and Guan C 2013 Brain-Computer Interface in Stroke Rehabilitation *Journal of Computer Science and Engineering*

[12] Pires G, Torres M, Casaleiro N, Nunes U and Castelo-Branco M 2011 Playing Tetris with non-invasive BCI *2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH)* 2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH) pp 1–6

[13] Reuderink B, Nijholt A and Poel M 2009 Affective Pacman: A Frustrating Game for Brain-Computer Interface Experiments *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering* vol 9 pp 221–7

[14] Krauledat M, Grzeska K, Sagebaum M, Blankertz B, Vidaurre C, Müller K-R and Schröder M 2009 Playing Pinball with non-invasive BCI *Advances in Neural Information Processing Systems 21* ed D Koller, D Schuurmans, Y Bengio and L Bottou (Curran Associates, Inc.) pp 1641–1648

[15] Laar B van de, Gürkök H, Bos D P-O, Poel M and Nijholt A 2013 Experiencing BCI Control in a Popular Computer Game *IEEE Transactions on Computational Intelligence and AI in Games* **5** 176–84

[16] He B, Baxter B, Edelman B J, Cline C C and Ye W W 2015 Noninvasive Brain-Computer Interfaces Based on Sensorimotor Rhythms *Proceedings of the IEEE* **103** 907–25

[17] Yang Y, Chevallier S, Wiart J and Bloch I 2012 Time-frequency selection in two bipolar channels for improving the classification of motor imagery EEG *Conf Proc IEEE Eng Med Biol Soc* **2012** 2744–7

[18] Zabidi A, Mansor W, Lee Y K and Fadzal C W N F C W 2012 Short-time Fourier Transform analysis of EEG signal generated during imagined writing *2012 International Conference on System Engineering and Technology (ICSET)*

2012 International Conference on System Engineering and Technology (ICSET) pp 1–4

[19] Gandhi T, Panigrahi B K and Anand S 2011 A comparative study of wavelet families for EEG signal classification *Neurocomputing* **74** 3051–7

[20] Amin H U, Malik A S, Ahmad R F, Badruddin N, Kamel N, Hussain M and Chooi W-T 2015 Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques *Australas Phys Eng Sci Med* **38** 139–49

[21] Prochazka A, Kukal J and Vysata O 2008 Wavelet transform use for feature extraction and EEG signal segments classification *2008 3rd International Symposium on Communications, Control and Signal Processing* 2008 3rd International Symposium on Communications, Control and Signal Processing pp 719–22

[22] Zhang X, Yao L, Huang C, Sheng Q and Wang X 2017 Intent Recognition in Smart Living Through Deep Recurrent Neural Networks *The 24th International Conference On Neural Information Processing (ICONIP 2017).*

[23] Zhang X, Yao L, Sheng Q Z, Kanhere S S, Gu T and Zhang D 2017 Converting Your Thoughts to Texts: Enabling Brain Typing via Deep Feature Learning of EEG Signals *arXiv:1709.08820 [cs]*

[24] Carvalho S R, Filho I C, Resende D O D, Siravenha A C, Souza C D, Debarba H G, Gomes B and Boulic R 2017 A Deep Learning Approach for Classification of Reaching Targets from EEG Images *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) pp 178–84

[25] An X, Kuang D, Guo X, Zhao Y and He L 2014 A Deep Learning Method for Classification of EEG Data Based on Motor Imagery *Intelligent Computing in Bioinformatics* Lecture Notes in Computer Science (Springer, Cham) pp 203–10

[26] Schirrmeister R T, Springenberg J T, Fiederer L D J, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W and Ball T 2017 Deep learning with convolutional neural networks for EEG decoding and visualization *Hum Brain Mapp* **38** 5391–420

[27] Lawhern V J, Solon A J, Waytowich N R, Gordon S M, Hung C P and Lance B J 2016 EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces *arXiv:1611.08024 [cs, q-bio, stat]*

[28] Bashivan P, Rish I, Yeasin M and Codella N 2015 Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks

[29] Lu N, Li T, Ren X and Miao H 2017 A Deep Learning Scheme for Motor Imagery Classification based on Restricted Boltzmann Machines *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25** 566–76

[30] Tabar Y R and Halici U 2017 A novel deep learning approach for classification of EEG motor imagery signals *J Neural Eng* **14** 016003

[31] Edelman B J, Baxter B and He B 2016 EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks *IEEE Transactions on Biomedical Engineering* **63** 4–14

[32] Luo J, Feng Z, Zhang J and Lu N 2016 Dynamic frequency feature selection based approach for classification of motor imageries *Comput. Biol. Med.* **75** 45–53

[33] Li M, Zhu W, Liu H and Yang J 2017 Adaptive Feature Extraction of Motor Imagery EEG with Optimal Wavelet Packets and SE-Isomap *Applied Sciences* **7** 390

[34] Saa J F D and Çetin M 2012 A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data *Journal of neural engineering* **9** 026020

[35] Ang K K, Chin Z Y, Zhang H and Guan C 2008 Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) pp 2390–7

[36] Yahya-Zoubir B, Bentlemsan M, Zemouri E-T and Ferroudji K 2015 Adaptive Time Window for EEG-based Motor Imagery Classification *Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication* IPAC '15 (New York, NY, USA: ACM) pp 83:1–83:6

[37] Anon 2012 Mutual information-based selection of optimal spatial–temporal patterns for single-trial EEG-based BCIs *Pattern Recognition* **45** 2137–44

[38] Sakhavi S, Guan C and Yan S 2015 Parallel convolutional-linear neural network for motor imagery classification *2015 23rd European Signal Processing Conference (EUSIPCO)* 2015 23rd European Signal Processing Conference (EUSIPCO) pp 2736–40

[39] Tayeb Z, Erçelik E and Conradt J 2017 Decoding of motor imagery movements from EEG signals using SpiNNaker neuromorphic hardware *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)* 2017 8th International IEEE/EMBS Conference on Neural Engineering (NER) pp 263–6

[40] Niranjani A N and Sivachitra M 2017 Motor imagery signal classification using spiking neural network *2017 International Conference on Intelligent Sustainable Systems (ICISS)* 2017 International Conference on Intelligent Sustainable Systems (ICISS) pp 901–4

[41] Yu Q, Tang H, Tan K C and Yu H 2014 A brain-inspired spiking neural network model with temporal encoding and learning *Neurocomputing* **138** 3–13

[42] Yu Q, Yan R, Tang H, Tan K C and Li H 2016 A Spiking Neural Network System for Robust Sequence Recognition *IEEE Transactions on Neural Networks and Learning Systems* **27** 621–35

[43] Zhang D, Yao L, Zhang X, Wang S, Chen W, Boots R and Benatallah B 2018 Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-based Intention Recognition for Brain Computer Interface *AAAI*

[44] Min S, Lee B and Yoon S 2016 Deep Learning in Bioinformatics *arXiv:1603.06430 [cs, q-bio]*

[45] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44

[46] Shahid S, Sinha R K and Prasad G 2010 Mu and beta rhythm modulations in motor imagery related post-stroke EEG: a study under BCI framework for post-stroke rehabilitation *BMC Neurosci* **11** P127

[47] McFarland D J, Miner L A, Vaughan T M and Wolpaw J R 2000 Mu and Beta Rhythm Topographies During Motor Imagery and Actual Movements *Brain Topogr* **12** 177–86

[48] Djemal R, Bazyed A G, Belwafi K, Gannouni S and Kaaniche W 2016 Three-Class EEG-Based Motor Imagery Classification Using Phase-Space Reconstruction Technique *Brain Sci* **6**

[49] Liu Y-H, Lin L-F, Chou C-W, Chang Y, Hsiao Y-T and Hsu W-C 2019 Analysis of Electroencephalography Event-Related Desynchronisation and Synchronisation Induced by Lower-Limb Stepping Motor Imagery *J. Med. Biol. Eng.* **39** 54–69

[50] Weber E and Doppelmayr M 2016 Kinesthetic motor imagery training modulates frontal midline theta during imagination of a dart throw *International Journal of Psychophysiology* **110** 137–45

[51] Clevert D-A, Unterthiner T and Hochreiter S 2015 Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) *arXiv:1511.07289 [cs]*

[52] Guennec A L, Malinowski S and Tavenard R 2016 Data Augmentation for Time Series Classification using Convolutional Neural Networks ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data

[53] Kim S K, Kirchner E A, Stefes A and Kirchner F 2017 Intrinsic interactive reinforcement learning – Using error-related potentials for real world human-robot interaction *Scientific Reports* **7** 17562

[54] Lotte F 2015 Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain #x2013;Computer Interfaces *Proceedings of the IEEE* **103** 871–90

[55] Wang F, Zhong S, Peng J, Jiang J and Liu Y 2018 Data Augmentation for EEG-Based Emotion Recognition with Deep Convolutional Neural Networks pp 82–93

[56] Li Y, Huang J, Zhou H and Zhong N 2017 Human Emotion Recognition with Electroencephalographic Multidimensional Features by Hybrid Deep Neural Networks *Applied Sciences* **7** 1060

[57] BCI Competition 2008 - Graz data sets 2A and 2B. *http://www.bbci.de/competition/iv/.*

[58] Klem G H, Lüders H O, Jasper H H and Elger C 1999 The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology *Electroencephalogr Clin Neurophysiol Suppl* **52** 3–6

[59] Shahid S and Prasad G 2011 Bispectrum-based feature extraction technique for devising a practical brain–computer interface *J. Neural Eng.* **8** 025014

[60] Ang K K, Chin Z Y, Wang C, Guan C and Zhang H 2012 Filter Bank Common Spatial Pattern Algorithm on BCI Competition IV Datasets 2a and 2b *Front Neurosci* **6**

[61] Zheng Q, Zhu F and Heng P 2018 Robust Support Matrix Machine for Single Trial EEG Classification *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26** 551–62

[62] Lotte F and Guan C 2011 Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms IEEE Transactions on Biomedical Engineering 58 355–62

[63] Raza H, Cecotti H, Li Y and Prasad G 2016 Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface Soft Comput 20 3085–96

[64] Gaur P, Pachori R B, Hui Wang and Prasad G 2015 An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain-computer interface 2015 International Joint Conference on Neural Networks (IJCNN) 2015 International Joint Conference on Neural Networks (IJCNN) pp 1–7

[65] Gaur P, Pachori R B, Wang H and Prasad G 2018 A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry Expert Systems with Applications 95 201–11